

## ABSTRACT OF THE DISCLOSURE

### CHINESE CHARACTER-BASED PARSER

5           A parser is provided that parses a Chinese text stream at the character level and  
builds a syntactic structure of Chinese character sequences. A character-based syntactic  
parse tree contains word boundaries, part-of-speech tags, and phrasal structure  
information. Syntactic knowledge constrains the system when it determines word  
boundaries. A deterministic procedure is used to convert word-based parse trees into  
10 character-based trees. Character-level tags are derived from word-level part-of-speech  
tags and word-boundary information is encoded with a positional tag. Word-level parts-  
of-speech become a constituent label in character-based trees. A maximum entropy  
parser is then built and tested.